

RESSALVA

Atendendo solicitação da autora,
o texto completo desta tese será
disponibilizado somente a partir
de 29/08/2021



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Camila Sthéfanie Colombo

Avaliação de Proficiência Oral em Língua Estrangeira:

Uma Proposta de Abordagem Avaliativa em um Exame para Professores de
Línguas com base em Métodos de Estruturação de Problemas e em Métodos
Multicritério de Análise da Decisão

São José do Rio Preto

2019

Camila Sthéfanie Colombo

Avaliação de Proficiência Oral em Língua Estrangeira:

Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Estudos Linguísticos, junto do Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Orientador: Prof. Dr. Douglas Altamiro Consolo.

São José do Rio Preto

2019

Colombo, Camila Sthéfanie.

C718a Avaliação de Proficiência Oral em Língua Estrangeira: Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com Base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão / Camila Sthéfanie Colombo. -- São José do Rio Preto, 2019
262 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Instituto de Biociências, Letras e Ciências Exatas, São José do Rio
Preto

Orientador: Douglas Altamiro Consolo

1. Linguística Aplicada. 2. Avaliação de Proficiência. 3.
Proficiência de Professores de Línguas. 4. Proficiência Oral. 5.
Inglês como Língua Estrangeira. I. Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências
Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Camila Sthéfanie Colombo

Avaliação de Proficiência Oral em Língua Estrangeira:

Uma Proposta de Abordagem Avaliativa em um Exame para Professores de Línguas com base em Métodos de Estruturação de Problemas e em Métodos Multicritério de Análise da Decisão

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Estudos Linguísticos, junto do Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. Douglas Altamiro Consolo
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Alexandre Bevilacqua Leoneti
USP – Ribeirão Preto

Profa. Dra. Vera Lúcia Teixeira da Silva
UFSCar – Câmpus São Carlos

Profa. Dra. Solange Aranha
UNESP – Câmpus de São José do Rio Preto

Profa. Dra. Suzi Marques Spatti Cavalari
UNESP – Câmpus de São José do Rio Preto

Suplentes

Profa. Dra. Teresa Helena Buscato Martins
UFBA – Câmpus Universitário Ondina

Profa. Dra. Paula Tavares Pinto
UNESP – Câmpus de São José do Rio Preto

Profa. Dra. Sandra Regina Butros Gattolin de Paula
UFSCar – Câmpus São Carlos

São José do Rio Preto
29 de agosto de 2019

Às mulheres da ciência brasileira.

AGRADECIMENTOS

A Deus, pela vida;

Ao meu orientador, Prof. Dr. Douglas Altamiro Consolo, pela oportunidade, pela confiança, pelo incentivo, pela compreensão e pelo exemplo de profissionalismo.

À Profa. Dra. Solange Aranha, à Profa. Dra. Suzi Marques Spatti Cavalari, à Profa. Dra. Vera Lúcia Teixeira da Silva e ao Prof. Dr. Alexandre Bevilacqua Leoneti pelas contribuições e pela disponibilidade para debater este trabalho.

À Profa. Dra. Maria Helena Vieira Abrahão, à Profa. Dra. Teresa Helena Buscato Martins, à Profa. Dra. Paula Tavares Pinto e à Profa. Dra. Sandra Regina Butros Gattolin de Paula, pela disponibilidade para compor o corpo suplente do Exame de Qualificação e da Defesa deste trabalho. Em especial, à Profa. Sandra, pelas reflexões tecidas na ocasião do IX SeLin.

Ao Prof. Dr. Edson Rosa Francisco de Souza pelos conhecimentos compartilhados na orientação de meu trabalho de qualificação especial.

À Profa. Dra. Cláudia Maria Ceneviva Nigro, por despertar meu interesse e oportunizar minha iniciação no universo da pesquisa.

A todos os docentes e funcionários do IBILCE e, em especial, aos associados ao Programa de Pós-Graduação em Estudos Linguísticos, pela qualidade dos serviços prestados e por garantirem um ambiente de formação acadêmica de qualidade reconhecida e mentalmente saudável.

Ao IFSP – Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, por incentivar e fomentar a qualificação de seu corpo docente. Em particular, a meus colegas do Câmpus Barretos, pelo suporte constante a meus compromissos de pesquisa.

Ao meu marido, Kaio, pelo amor, pelo companheirismo, pelo suporte, pela inspiração e pelos incentivos constantes.

Aos meus pais, Aparecida e João, e a meu irmão, João Victor, por todo incentivo e carinho.

A minha sogra, Cecília, pelo apoio e pelo carinho.

A meus avós, Aparecida e Antônio, Pedro e Elydia (*in memorian*), pelo carinho.

Aos membros do grupo de pesquisa ENAPLE-CCC, Aline, Elen, Rodrigo, Jéssica, William, Diego e José Ricardo, pelos conhecimentos compartilhados.

A minhas colegas de trabalho Aline, Marcela, Natalia, Pricila, Rita, Rozângela, Laura e Isabel, pela compreensão, pelo apoio e pela descontração. Em especial, à Marcela, pela amizade.

Às amigas Jaqueline, Michele, Adelaide e Carollina, pela amizade, pela torcida e pelo incentivo.

À Daniela Bianchi Pandim, Camila Uenoyama Moura Miranda Pereira e Mariana dos Santos Murra, pelos cuidados propiciados durante esta caminhada.

Aos participantes desta pesquisa, que, voluntariamente, dedicaram parte de seu tempo ao compartilhamento de conhecimento.

A todos aqueles que se dedicam ao desenvolvimento da pesquisa brasileira e à manutenção das garantias de acesso ao ensino público gratuito.

A todos aqueles que, direta ou indiretamente, contribuíram para o desenvolvimento desta pesquisa.

Eu vi um menino correndo
Eu vi o tempo brincando ao redor
Do caminho daquele menino

Eu pus os meus pés no riacho
E acho que nunca os tirei
O sol ainda brilha na estrada e eu nunca passei

Eu vi a mulher preparando outra pessoa
O tempo parou pra eu olhar para aquela barriga
A vida é amiga da arte
É a parte que o sol me ensinou
O sol que atravessa essa estrada que nunca passou
[...]

Força Estranha, Caetano Veloso (1978)

RESUMO

Neste trabalho investiga-se a abordagem de avaliação da proficiência oral do professor de línguas estrangeiras empregado por um exame brasileiro voltado a essa finalidade específica, o EPPL – Exame de Proficiência para Professores de Línguas Estrangeiras. O estudo tem por objetivo propor uma abordagem que atenda a necessidades levantadas pela literatura da área acerca da importância da democratização e transparência do processo avaliativo e do emprego de processos decisórios que garantam os princípios de validade sem causar interferências na praticidade das avaliações. Assim, a abordagem proposta subordina-se à elaboração de uma escala de mensuração e à reestruturação da escala de proficiência do exame, agregando-se a opinião de usuários com diferentes graus de especialidade sobre a avaliação de proficiência. Trata-se, portanto, de um estudo de caso de natureza híbrida e cunho exploratório-descritivo que se utiliza de dados secundários e primários para auxiliar a investigação. Os dados secundários consistem em estudos e documentos sobre a caracterização da proficiência linguística para finalidades gerais e para finalidades de ensino, bem como sobre a avaliação dessas proficiências; em dados de fala do teste oral do EPPL referentes à participação de formandos em Letras de uma universidade pública paulista e de duas universidades públicas mineiras, na qualidade de candidatos ao exame, nos anos de 2015 e 2017; e nos respectivos conceitos atribuídos aos candidatos, seguindo-se os princípios determinados pela escala de proficiência oral original do exame. Os dados primários, por sua vez, consistem nas respostas obtidas por meio da aplicação sucessiva de dois questionários estruturados com base nos princípios do método de estruturação de problemas VFT – Pensamento Focado no Valor; o primeiro, com a finalidade de coletar informações a respeito do perfil dos participantes da pesquisa e de suas opiniões sobre as características da proficiência do professor de línguas e de sua avaliação e o segundo, para refinar e confirmar a opinião dos participantes, fornecendo embasamento para a elaboração da escala de mensuração. A determinação dos critérios que integram a escala de mensuração proposta, bem como de seus pesos para a avaliação, foi realizada com base nos valores informados pelos usuários do exame, nos valores determinados pela literatura e na criterialidade encontrada por meio da análise das características de fala dos candidatos. Os perfis linguísticos que representam os níveis de variação da proficiência na escala de mensuração, por sua vez, foram definidos com base nos resultados fornecidos pela análise multivariada de agrupamento, mais especificamente, pelo método de Ward, desses desempenhos. A escala de proficiência foi elaborada considerando-se os quadros de referência internacional CEFR – Quadro Europeu Comum de Referência para Línguas e CETF – *Cambridge English Teaching Framework*. A reavaliação dos desempenhos de fala foi realizada considerando-se a escala de mensuração proposta combinada ao método multicritério de análise da decisão *Fuzzy TOPSIS* – Técnica de Ordenamento de Preferências por Similaridade à Solução Ideal. A comparação do nível de proficiência sugerido pela aplicação do método em relação à escala de proficiência proposta com base nos quadros de referência mostrou-se confiável e, em relação à avaliação original, permitiu o tratamento de discrepâncias.

Palavras-chave: Avaliação. Proficiência oral. Proficiência do professor de línguas. Métodos de estruturação de problemas. Métodos multicritério de análise da decisão.

ABSTRACT

This research study investigates the assessment model used by a teaching-oriented language examination, the EPPLÉ – Proficiency Examination for Foreign Language Teachers, in the assessment of foreign language teachers' oral proficiency. We aim at evaluating the model used by the oral test and proposing a new one that meets current concerns raised by the literature regarding the importance of democratizing the assessment process, guaranteeing transparency and using decision models that assure validity principles without causing interferences in the test's practicality. Therefore, the model we propose relies on the reorganization of the oral proficiency scale used in the exam through the inclusion of its stakeholders' opinions. The investigation conducted consists on a quali-quantitative and exploratory-descriptive case study that uses primary and secondary data. Secondary data consist of research studies and documents on the concept of general and teaching purpose language proficiency; speaking data from candidates to the oral test of the EPPLÉ examination; as well as the respective rates assigned to such candidates according to the proficiency scale used in the exam. The candidates to the EPPLÉ examination that participate in this study are Letters course undergraduate students from Brazilian public universities, one located in the state of São Paulo and two in the state of Minas Gerais, in the years of 2015 and 2017. Primary data, on the other hand, consist of the answers obtained through the application of two questionnaires, which we structured following the principles of a problem structuring method, VFT – Value-Focused Thinking. We applied the first questionnaire aiming at collecting information on the research study participants' profiles and opinions concerning foreign language teachers' proficiency and its assessment. The second questionnaire supported the reorganization of the scale and was used to refine and confirm the participants' opinions. We selected the criteria included in the rubric scale, as well as its assessment weights, through the values of the stakeholders, the values pointed out in the literature and their criteriality, determined through the analysis of candidates' speech samples. Language profiles representing the variation of the levels of the proficiency in the rubric were defined through performance cluster analysis, more specifically through Ward's method. The proficiency scale was created through the principles presented in the CEFR – Common European Framework of Reference for Languages and in the CETF – Cambridge English Teaching Framework. Finally, the reassessment of the performances was conducted through the combination of the rubric and the multi-criteria decision analysis method, Fuzzy TOPSIS – Technique for Order of Preference by Similarity to Ideal Solution. The comparison of the proficiency level suggested by the application of the method based on the proficiency scale resulted in mainly reliable results, allowing the management of error measurements.

Keywords: Language testing. Oral proficiency. Problem Structuring Methods. Multicriteria Decision-Analysis methods. Teachers' oral proficiency

LISTA DE ILUSTRAÇÕES

Figuras

Figura 1. Ligações entre argumentos interpretativos.	47
Figura 2. Validade de construto da interpretação dos resultados.	48
Figura 3. Interação entre construto, critérios e o teste.	50
Figura 4. Etapas do processo de análise da decisão.	80
Figura 5. Modelo de matriz de decisão	85
Figura 6. Passos para aplicação do método TOPSIS clássico.	86
Figura 7. Passos para aplicação do método Fuzzy TOPSIS proposto por Chen (2000).	89
Figura 8. Definição dos pesos agregados e do número de critérios e de decisores.	108
Figura 9. Avaliação dos critérios em relação às alternativas.	108
Figura 10. Resultado da classificação sugerida.	108
Figura 11. Triangulação dos dados do estudo.	110
Figura 12. Competências e critérios relevantes para a atuação do professor de línguas segundo usuários do exame.	117
Figura 13. Competências e critérios relevantes para a atuação do professor de línguas segundo os envolvidos com a elaboração do exame.	120
Figura 14. Tarefa de apresentação pessoal da versão 3.0 do teste oral do exame EPPL.	143
Figura 15. Tarefa de apresentação pessoal da versão 3.1 do teste oral do exame EPPL.	143
Figura 16. Tarefa descrição de figuras da versão 3.1 do teste oral do exame EPPL.	150
Figura 17. Tarefa de discussão de tópico da versão 3.0 do teste oral do exame EPPL.	157
Figura 18. Tarefa de discussão de tópico da versão 3.1 do teste oral do exame EPPL.	158
Figura 19. Rubrica da Tarefa 4 do teste oral do exame EPPL (Versão 3.1).	163
Figura 20. Tarefa de metalinguagem da versão 3.0.1 do teste oral do exame EPPL.	170
Figura 21. Tarefa de metalinguagem da versão 3.1 do teste oral do exame EPPL.	171
Figura 22. Relação entre os componentes da competência linguístico-comunicativa avaliados.	180
Figura 23. Acento nuclear de DES.	183
Figura 24. Acento nuclear de ROB.	183

Quadros

Quadro 1. Atividades linguístico-comunicativas do CEFR.	56
Quadro 2. Estratégias linguístico-comunicativas do CEFR e criterialidade.	57
Quadro 3. Competências linguístico-comunicativas do CEFR e criterialidade.	58
Quadro 4. Criterialidade e tipos de erros gramaticais do CEFR.	60

Quadro 5. Quadro-resumo das principais características avaliativas dos exames internacionais.....	67
Quadro 6. Resultados dos exames em relação ao CEFR.....	68
Quadro 7. Etapas de construção da rede de objetivos meios e da hierarquia de objetivos fundamentais.....	83
Quadro 8. Quadro-resumo dos dados de fala.....	101
Quadro 9. Códigos empregado nas transcrições.....	104
Quadro 10. Relação entre as competências e os critérios do CEFR e do CETF e a opinião dos participantes.....	115
Quadro 11. Relação entre as competências e os critérios do CEFR e do CETF e a opinião dos participantes.....	118
Quadro 12. Resumo da importância dos critérios na opinião dos stakeholders.....	121
Quadro 13. Hierarquização dos objetivos do exame EPPL.....	124
Quadro 14. Variáveis de uso e conhecimento sobre regras linguísticas da escala do exame EPPL.....	135
Quadro 15. Hierarquização dos objetivos fundamentais segundo os princípios do VFT.....	136
Quadro 16. Definição dos atributos/critérios de mensuração.....	136
Quadro 17. Limite superior e inferior do atributo “clareza” para o objetivo “avaliar o fornecimento de informações sobre experiências”.....	137
Quadro 18. Limite superior e inferior dos atributos “velocidade” e “ritmo” para o objetivo “avaliar a fluência”.....	137
Quadro 19. Limite superior e inferior dos atributos “inteligibilidade”, “compreensibilidade” e “proximidade” para o objetivo “avaliar a pronúncia”.....	137
Quadro 20. Limite superior e inferior dos atributos “acurácia” e “complexidade” para o objetivo “avaliar o uso de estruturas gramaticais”.....	137
Quadro 21. Limite superior e inferior dos atributos “acurácia” e “variedade” para o objetivo “avaliar o uso de estruturas lexicais”.....	138
Quadro 22. Limite superior e inferior do atributo “adequação” para o objetivo “avaliar a expressão de conhecimento sobre regras linguísticas”.....	138
Quadro 23. Limite superior e inferior do atributo “clareza” para o objetivo “avaliar a seleção de informações para esclarecer dúvidas linguísticas”.....	138
Quadro 24. Limite superior e inferior dos atributos “adequação” e “compreensibilidade” para o objetivo “avaliar o fornecimento de explicação sobre regras linguísticas”.....	138
Quadro 25. Limite superior e inferior do atributo “adequação” para o objetivo “avaliar o uso de terminologia específica”.....	138
Quadro 26. Especificações das tarefas do teste oral do exame EPPL (Versão 3.0 e 3.1).....	141
Quadro 27. Recursos terminológicos empregados no texto instrutivo.....	169

Quadro 28. Terminologia empregada na Situação 1 da Tarefa 3 do teste oral do EPPLÉ.....	178
Quadro 29. Terminologia empregada na Situação 2 da Tarefa 3 do teste oral do EPPLÉ.....	178
Quadro 30. Terminologia empregada na Situação A da Tarefa 5 do teste oral do EPPLÉ.....	178
Quadro 31. Terminologia empregada na Situação B da Tarefa 5 do teste oral do EPPLÉ.....	179
Quadro 32. Escala de mensuração para o teste oral do EPPLÉ.	197
Quadro 33. Escala de proficiência oral sugerida ao exame EPPLÉ com base em quadros de referências internacionais.....	200
Quadro 34. CEFR em relação a exames consagrados e ao EPPLÉ.....	204
Quadro 35. Definição dos pesos dos critérios de avaliação do teste oral do EPPLÉ.....	205
Quadro 36. Variáveis linguísticas associadas às faixas de proficiência.....	206
Quadro 37. Seleção de candidatos para reavaliação.	208
Quadro 38. Perfil de desempenho de LID.	209
Quadro 39. Perfil de desempenho de THY.	210
Quadro 40. Perfil de desempenho de DES.	212
Quadro 41. Perfil de desempenho de LIG.	212
Quadro 42. Perfil de desempenho de MAC.	212
Quadro 43. Perfil de desempenho de ROB.....	212
Quadro 44. Perfil de desempenho de LAC.....	212
Quadro 45. Quadro-resumo dos resultados da avaliação original e da reavaliação dos desempenhos de fala.	214

Gráficos

Gráfico 1. Competências da avaliação da proficiência do professor de línguas na opinião dos usuários do exame.....	123
Gráfico 2. Dendrograma da quantidade de funções comunicativas utilizadas na tarefa de apresentação pessoal.	146
Gráfico 3. Dendrograma da quantidade de emprego de tópicos expandidos na tarefa de apresentação pessoal.	147
Gráfico 4. Dendrograma do emprego de funções comunicativas na tarefa de descrição de figuras.....	153
Gráfico 5. Dendrograma da frequência do emprego de função comunicativa por nível de proficiência.....	154
Gráfico 6. Dendrograma do emprego de funções comunicativas na tarefa de discussão de tópico.....	160
Gráfico 7. Dendrograma da frequência de extensão de tópico na tarefa de discussão.....	161
Gráfico 8. Dendrograma do emprego de funções no fornecimento de instruções.....	166
Gráfico 9. Dendrograma do emprego de recurso terminológico na tarefa de instrução.....	168

Gráfico 10. Dendrograma do emprego de recursos metalinguísticos.....	174
Gráfico 11. Dendrograma da porcentagem de acurácia da explicação metalinguística.....	176
Gráfico 12. Dendrograma do desempenho de fala com relação ao acento nuclear.....	182
Gráfico 13. Dendrograma dos desempenhos de fala com relação à velocidade de fala.....	184
Gráfico 14. Dendrograma dos desempenhos de fala com relação ao alongamento de sons.....	185
Gráfico 15. Dendrograma dos desempenhos de fala com relação às pausas silenciosas.....	186
Gráfico 16. Dendrograma dos desempenhos de fala com relação às pausas preenchidas.....	187
Gráfico 17. Dendrograma dos desempenhos de fala com relação às repetições.....	188
Gráfico 18. Dendrograma dos desempenhos de fala com relação às reformulações.....	189
Gráfico 19. Dendrograma dos desempenhos de fala com relação à precisão linguística.....	190
Gráfico 20. Dendrograma dos desempenhos de fala com relação à medida global.....	191
Gráfico 21. Dendrograma dos desempenhos de fala com relação ao índice de coordenação.....	191
Gráfico 22. Dendrograma dos desempenhos de fala com relação ao Índice de Guiraud.....	192
Gráfico 23. Dendrograma dos desempenhos de fala com relação à frequência de palavras.....	193

LISTA DE TABELAS

Tabela 1. Tabela de contingência de conhecimento e participação no exame EPPL.....	112
Tabela 2. Resumo descritivo do emprego mediano da quantidade de funções comunicativas abordadas.....	146
Tabela 3. Resumo descritivo do emprego mediano de tópicos expandidos.....	148
Tabela 4. Resumo descritivo do emprego mediano de funções comunicativas para descrição.....	153
Tabela 5. Resumo descritivo do emprego mediano de funções comunicativas para opinião.....	160
Tabela 6. Resumo descritivo do emprego mediano de organização do texto instrutivo.....	166
Tabela 7. Resumo descritivo do emprego mediano de variedade terminológica no texto instrutivo.....	168
Tabela 8. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia.....	174
Tabela 9. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia.....	177
Tabela 10. Resumo descritivo do emprego mediano de explicação metalinguística com acurácia.....	177
Tabela 11. Frequência do emprego de acento nuclear.....	182
Tabela 12. Resumo descritivo do emprego mediano de velocidade de fala.....	184
Tabela 13. Resumo descritivo do emprego mediano de alongamento de sons.....	185
Tabela 14. Resumo descritivo do emprego mediano de pausas silenciosas.....	186
Tabela 15. Resumo descritivo do emprego mediano de pausas preenchidas.....	187
Tabela 16. Resumo descritivo do emprego mediano de repetições.....	188
Tabela 17. Resumo descritivo do emprego mediano de reparos.....	189
Tabela 18. Resumo descritivo do emprego mediano de precisão linguística.....	190
Tabela 19. Resumo descritivo do emprego mediano de medida global.....	191
Tabela 20. Resumo descritivo do emprego mediano do índice de coordenação.....	192
Tabela 21. Resumo descritivo do emprego mediano de Índice de Guiraud.....	193
Tabela 22. Resumo descritivo do emprego mediano de frequência de palavras.....	194
Tabela 23. Mediana ou acima do limiar de A.....	207
Tabela 24. Limiar inferior de A.....	207
Tabela 25. Limiar superior de B.....	207
Tabela 26. Mediana de B.....	207
Tabela 27. Limiar inferior de B.....	207
Tabela 28. Limiar superior de C.....	207
Tabela 29. Mediana de C.....	207
Tabela 30. Limiar inferior de C.....	207
Tabela 31. Limiar superior de D.....	207
Tabela 32. Mediana ou inferior ao limiar de D.....	207
Tabela 33. Desempenho insuficiente.....	207

LISTA DE ABREVIACÕES

AFT – *Alternative Focused Thinking*

AHP – *Analytic Hierarchy Process*

CAE – *Certificate of Advanced English*

CEFR – *Common European Framework of Reference for Languages*

CETF – *Cambridge English Teaching Framework*

CLAN – *Computerized Language Analysis*

CLC – Competência Linguístico-Comunicativa

CPE – *Certificate of Proficiency in English*

DT – *Drama Theory*

EPPL – Exame de Proficiência para Professores de Língua Estrangeira

ETS – *English Testing Service*

FCE – *First Certificate in English*

IELTS – *International English Language Testing System*

KET – *Key English Test*

LE – Língua(s) Estrangeira(s)

MCDA – *Multicriteria Decision Analysis* (Método Multicritério de Análise da Decisão)

PLCPP – Proficiência Linguístico-Comunicativo-Pedagógica

PSMs – *Problem Structuring Methods* (Métodos de Estruturação de Problemas)

SCA – *Strategic Choice Approach*

SODA – *Strategic Options Development and Analysis*

SSM – *Soft Systems Method*

TEPOLI – Teste de Proficiência Oral em Língua Inglesa

TKT – *Teaching Knowledge Test*

TOEFL iBT – *Test of English as a Foreign Language*

TOPSIS – *Technique for Order of Preference by Similarity to Ideal Solution*

VFT – *Value Focused Thinking*

SUMÁRIO

1 INTRODUÇÃO.....	25
1.1 A pesquisa	29
1.2 Organização da tese	33
2 CAPÍTULO I: FUNDAMENTAÇÃO TEÓRICA	35
2.1 Avaliação de proficiência de línguas: conceitos iniciais e desafios	35
2.1.1 Percurso histórico dos modos de avaliação da proficiência linguística: da era pré-científica à contemporaneidade	39
2.1.2 O desenvolvimento de exames de proficiência linguística: o modelo de utilidade do teste e suas qualidades	43
2.1.2.1 A avaliação da produção oral em língua estrangeira: o Quadro Comum de Referências para Línguas (CEFR) e os critérios de exames de proficiência consagrados	53
2.2 Avaliação da proficiência do professor de línguas estrangeiras: competências e perspectivas	68
2.2.1 O tratamento da linguagem pedagógica no Cambridge English Teaching Framework no exame de formação de professores de línguas TKT	73
2.3 A ciência da decisão: da neurociência à neuroeconomia	74
2.3.1 Estruturação de problemas e análise da decisão	77
2.3.1.1 Métodos de Estruturação de Problemas (PSMs)	81
2.3.1.2 Métodos Multicritério de Análise da Decisão (MCDA)	84
2.3.1.2.1 Técnica para Avaliar o Desempenho de Alternativas através de Similaridade com a Solução Ideal: TOPSIS e Fuzzy TOPSIS	85
2.4 A decisão na avaliação de proficiência em Linguística Aplicada: a possibilidade de tratamento interdisciplinar das discrepâncias e da incerteza	91
2.5 O capítulo em resumo.....	95
3 CAPÍTULO II: METODOLOGIA DA PESQUISA	97
3.1 Natureza da pesquisa	97
3.2 Contexto e participantes da pesquisa.....	98
3.2.1 O teste oral do exame EPPLÉ	98
3.2.2 Os candidatos ao exame EPPLÉ.....	100
3.2.3 Os usuários (<i>stakeholders</i>) especialistas do exame	102
3.3 Procedimentos e instrumentos de coleta e geração de dados	103
3.3.1 Transcrições das gravações em áudio do desempenho de fala de candidatos ao teste oral do exame EPPLÉ	103
3.3.2 Questionários para usuários (<i>stakeholders</i>) especialistas do exame EPPLÉ.....	105
3.4 Procedimentos e instrumentos de análise e triangulação de dados	105
3.5 O capítulo em resumo.....	110

4 CAPÍTULO III: DISCUSSÃO E ANÁLISE DE DADOS	111
4.1 Em busca de critérios para avaliação da proficiência oral do professor de línguas: os interesses dos usuários (<i>stakeholders</i>) do exame	111
4.1.1 As respostas fornecidas ao questionário de levantamento de critérios	111
4.1.2 As respostas fornecidas aos questionário sobre mensuração da proficiência	121
4.1.3 A estruturação do problema: considerações gerais sobre a opinião dos usuários.....	124
4.2 Avaliação da proficiência do professor de línguas pelo EPPL	125
4.2.1 O construto informado e representado do exame: hierarquização dos objetivos da escala de proficiência oral.....	125
4.2.2 Definição dos critérios de avaliação do desempenho oral	135
4.2.3 Avaliação da proficiência oral no EPPL: o desempenho dos candidatos nas tarefas de produção oral.....	139
4.2.3.1 Avaliação do uso da linguagem com finalidade geral no exame EPPL	139
4.2.3.1.1 A tarefa de apresentação pessoal do teste oral do exame EPPL.....	143
4.2.3.1.2 Avaliação do desempenho na apresentação pessoal no teste oral do exame EPPL	145
4.2.3.1.3 A tarefa de descrição de figuras do exame EPPL.....	150
4.2.3.1.4 Avaliação do desempenho na descrição de figuras no teste oral do exame EPPL	152
4.2.3.1.5 A tarefa de discussão de tópico do exame EPPL.....	157
4.2.3.1.6 Avaliação do desempenho na discussão de tópico no teste oral do exame EPPL	159
4.2.3.2 Avaliação do uso da linguagem com finalidade pedagógica no exame EPPL	162
4.2.3.2.1 A tarefa de fornecimento de instruções no teste oral do exame EPPL	163
4.2.3.2.2 Avaliação do desempenho no fornecimento de instruções no teste oral do exame EPPL.....	165
4.2.3.2.3 A tarefa de explicação gramatical do EPPL	169
4.2.3.2.4 Avaliação do desempenho na explicação gramatical pela escala de proficiência oral do exame EPPL	172
4.2.3.3 Avaliação das competências linguístico-comunicativas no exame EPPL.....	180
4.3 Proposição de uma abordagem avaliativa da proficiência oral do professor de línguas para o EPPL.....	194
4.3.1 A proposição de escalas para o teste oral do exame EPPL	194
4.3.1.1 A escala de mensuração analítica para o teste oral do exame EPPL	195
4.3.1.2 A escala de proficiência oral para o exame EPPL.....	198
4.3.2 A avaliação da proficiência oral com auxílio do método multicritério de análise da decisão <i>Fuzzy TOPSIS</i>	204
4.3.3 Os resultados da reavaliação dos desempenhos de fala	208
4.4 O capítulo em resumo	215

5 CONSIDERAÇÕES FINAIS	217
REFERÊNCIAS	221
APÊNDICES	234
APÊNDICE A – Características distintivas da escala de proficiência oral do EPPL	234
APÊNDICE B - Questionário de levantamento de interesses sobre o exame EPPL I.....	236
APÊNDICE C - Questionário de levantamento de interesses sobre o exame EPPL II.....	245
ANEXOS	259
ANEXO A - Escala Global do CEFR.....	259
ANEXO B - A escala do IELTS.....	260
ANEXO C - Escala de proficiência oral do EPPL	261

1 INTRODUÇÃO

O ato de avaliar consiste em uma característica historicamente universal da vida em sociedade (McNAMARA, 2000) e, sendo uma atividade inerente ao ser humano, permeia o cotidiano em diversos níveis de complexidade (MESSICK, 1994; SHOHAMY, 2001). Na Linguística Aplicada, a avaliação é entendida como uma construção social (McNAMARA; ROEVER, 2006), apresentando, basicamente, três grandes propósitos: avaliação do rendimento escolar de indivíduos (*assessment*), testagem de capacidades (*testing*) e avaliação de qualidade educacional (*evaluation*) (FULCHER; DAVIDSON, 2007). A avaliação escolar tem finalidades formativas e seus resultados são provenientes da interação existente entre os elementos do tripé educacional ensino-aprendizagem-avaliação ao longo do processo de instrução. Basicamente, essa forma de avaliar consiste na coleta e análise de informações a respeito do rendimento do aprendizado de alunos por meio de instrumentos de avaliação diversos, como provas, portfólios, avaliação entre pares, autoavaliação, dentre outros (GENESEE, 2001). A testagem, por sua vez, não tem por foco o processo instrutivo e, portanto, preocupa-se apenas com os resultados gerados pelo próprio instrumento de avaliação empregado. Desse modo, por meio da identificação e coleta de informações consideradas relevantes, visa a uma padronização das potenciais capacidades do examinado em relação a critérios previamente estabelecidos ou aos demais examinados (McNAMARA, 2000). Seus resultados, contudo, geram impactos que podem promover alterações relevantes nos cenários educacionais e sociais. A avaliação de proficiência linguística, objeto de estudo deste trabalho, consiste em um exemplo representante da categoria de testagem. Por fim, a avaliação de qualidade apoia-se em padrões estabelecidos para o processo educacional e pode ser empregada com a finalidade de prestação de contas sobre o processo instrucional; de nivelamento, promoção ou delimitação do status de um indivíduo em um curso ou programa; ou de orientação da instrução em sala de aula e consequente melhoria do processo de aprendizagem (GENESEE, 2001).

Independentemente do propósito específico da avaliação, o ato de avaliar em si consiste em uma estratégia que visa à coleta, análise e interpretação de informações consideradas relevantes que possibilitem a realização de inferências para tomadas de decisão apropriadas (McNAMARA, 2000; GENESEE, 2001). Na avaliação de proficiência linguística cabem aos avaliadores decidirem sobre os graus de adequação, representatividade e acurácia das amostras de língua fornecidas pelos candidatos e, por extensão, sobre a melhor faixa de proficiência que representa o potencial de uso da língua desse candidato com base no desempenho demonstrado

(McNAMARA, 2004); e, aos demais usuários (*stakeholders*) – como o próprio candidato, pesquisadores, líderes educacionais, formuladores de políticas educacionais, dentre outros –, as implicações sociais dos resultados apresentados (ELDER *et al.*, 2017). Assim, é importante ressaltar que o ato de avaliar, com seus propósitos e modelos, inevitavelmente, pressupõe tanto na etapa de avaliação *per se* quanto no período de pós-avaliação – aquele envolvendo o uso, pelos *stakeholders*, dos resultados fornecidos pelos examinadores – o ato de tomar decisões (KANE, 2002).

A decisão corresponde a uma atividade de posicionar-se em relação ao futuro (GOMES; GOMES; ALMEIDA, 2009), considerando-se o entendimento que se tem dessas projeções em um determinado momento político e sócio-histórico em relação a um ou mais critérios (BUEDE, 2013). No entanto, Munda (1993) destaca que sistemas sociais – aqueles nos quais os seres humanos têm um papel central – são marcados por imprecisão e incerteza devido a seu grande número de variáveis e de circuitos de retorno. Nesse sentido, quanto maiores os níveis de complexidade de um sistema, mais os conceitos de precisão e significância dos julgamentos realizados acerca de um comportamento tornam-se mutuamente excludentes (ZADEH, 1975), dificultando a tomada de decisões.

Para Clemen e Reilly (2013), a dificuldade inerente a algumas decisões está intimamente relacionada ao grau de singularidade do problema para o qual se busca uma solução. O problema consiste em uma oportunidade de decisão e estabelece-se quando se tem um objetivo em mente, mas não se sabe como atingi-lo (FISCHER; GREIFF; FUNKE, 2012). Desse modo, inicialmente, poder-se-ia pensar que a experiência com o tipo de problema seria suficiente para resolver a questão. No entanto, Clemen e Reilly (2013) ressaltam que, na verdade, o conhecimento proporcionado pela experiência acaba apenas por permitir o reconhecimento da dificuldade em problemas similares e a compreensão de como o problema consegue se encaixar em um cenário mais amplo para que a escolha das ferramentas apropriadas possa ser realizada. A dificuldade, desse modo, consiste não apenas na seleção dos elementos que compõem o processo, mas na tentativa de garantir que todos os elementos importantes sejam analisados. Assim, os autores destacam quatro fatores como potenciais geradores de dificuldade decisória: a complexidade do objeto, a incerteza sobre as informações, os múltiplos objetivos da decisão e os pontos de vista conflitantes por parte dos decisores.

A complexidade estabelece-se na relação de interdependência existente entre diversas características que compõem um objeto, implicando uma dificuldade de definição e consideração apropriada de todos os componentes críticos que integram problemas autóctones na análise da decisão diante da presença constante de elementos de ordem dinâmica e difusa

(MUNDA, 1993; CLEMEN; REILLY, 2013). A incerteza, por sua vez, consiste na ausência de informações desejadas para a decisão e, quando reflete probabilidades já conhecidas, por não permitir que o decisor visualize como os fatores que não estão sob seu controle irão se comportar, colocam a efetividade do processo decisório em risco (SMITH; HUETTEL, 2010; CLEMEN; REILLY, 2013). O trabalho com múltiplos objetivos, ante à limitação cognitiva dos seres humanos, pode fazer com que o foco em um objetivo específico acabe por impedir o alcance dos demais (CLEMEN; REILLY, 2013) e os pontos de vista conflitantes, por fim, podem ocorrer diante de alterações no insumo fornecido à compreensão do problema ou, simplesmente, por discordância entre os decisores a respeito dos valores subjetivos dos componentes, incluindo-se, aqui, fatores de interação social, como cooperativismo e honestidade (SMITH; HUETTEL, 2010; CLEMEN; REILLY, 2013). Por subjetividade, entenda-se a importância garantida aos critérios por cada um dos decisores envolvidos no processo, a qual é influenciada, especialmente, pelos graus de incerteza e de imprecisão de dados sobre as alternativas disponíveis (WĄTRÓBSKI *et al.*, 2018).

Na avaliação de línguas, a intenção de uma decisão deve ser a de maximizar o ideal de uma sociedade democrática (FULCHER; DAVIDSON, 2007). Assim, a avaliação de proficiência linguística é considerada uma atividade relevante em termos políticos, sociais e educacionais para a sociedade contemporânea, o que faz com que a Linguística Aplicada se preocupe constantemente com a clareza, confiabilidade e atualidade das informações nela e por meio dela veiculadas (SHOHAMY; OR, 2017). Tendo em vista que o objetivo da avaliação de proficiência linguística é realizar inferências acerca dos potenciais desempenhos que o candidato será capaz de realizar na vida real com base em amostras de desempenho que podem ser observadas em um exame ou teste (McNAMARA, 2000), entende-se que o verdadeiro objeto de interesse da avaliação de proficiência linguística consiste em um comportamento que não pode ser diretamente observado pelo avaliador, apenas inferido por meio dos construtos que embasam o referido instrumento (McNAMARA, 2004).

Diante disso, McNamara (2000) ressalta que o processo avaliativo é naturalmente limitado e, portanto, sujeito a erros, seja por questões de ordem contextual ou interpretativa, ou, ainda, por fatores afetivos gerados pela interferência do ato de observar no desempenho do candidato. Faz-se necessário, portanto, que a autenticidade dos testes e, por extensão, a modelagem realizada do construto sejam gerenciadas, de modo a minimizar interferências na possibilidade de generalização dos resultados (McNAMARA, 2004). Assim, é imprescindível ressaltar a inerência da subjetividade e da incerteza na avaliação de testes com objetos complexos, como é o caso da avaliação de proficiência linguística. A complexidade do objeto

acaba dificultando a capacidade dos elaboradores de delimitarem a dinamicidade interacional de seus componentes, fazendo com que os critérios sejam de difícil mensuração e, portanto, podendo gerar discrepâncias (BACHMAN; LYNCH; MASON, 1995). Ou seja, os resultados de uma avaliação de proficiência estão sujeitos a discrepâncias por configurarem uma decisão tomada subjetivamente com base em amostras de língua, as quais consistem em variáveis complexas e difusas, elicitadas por meio de respostas abertas sobre as quais são realizados inferências e julgamentos mais ou menos influenciados pelas opiniões, experiências e percepções do decisor a respeito do construto do exame e da amostra (HAMP-LYONS, 1992; ANG-AW; GOH, 2011).

Discrepâncias consistem em variabilidades no julgamento da qualidade do desempenho do candidato entre os avaliadores e, geralmente, são geradas pelo nível de severidade ou leniência desses avaliadores, pelo tempo de experiência com ensino de línguas e avaliação que eles apresentam, pelo grau de consistência no entendimento e emprego da escala solicitada, pela eficácia do treinamento realizado, dentre outros fatores (ANG-AW; GOH, 2011; TOFFOLI; SIMON, 2018). Ademais, Pollitt e Murray (1996) e Brown (2000) ressaltam que os avaliadores costumam diferir com relação às próprias abordagens avaliativas que empregam. De maneira geral, esses avaliadores podem formar um conceito holístico nos primeiros minutos acerca dos avaliados de modo que, na sequência, passem a buscar evidências de desempenho que comprovem ou refutem suas impressões iniciais.

Assim, apesar de serem inerentes ao processo decisório, especialmente na presença de múltiplos critérios e múltiplos decisores, as discrepâncias podem inviabilizar a obtenção dos resultados da avaliação ou, ainda, gerar um emprego desnecessário de recursos para sua solução. Buede (2013) destaca que o processo de tomar decisões envolve o dispêndio e a alocação de recursos visando a uma busca por mudança ou manutenção de um *status quo*. Independentemente da natureza desses recursos – financeira, temporal, cognitiva, dentre outros –, faz-se necessário apontar que um pensamento mais estruturado a respeito do problema decisório é imprescindível para sua melhor compreensão e, conseqüentemente, para a economia dos recursos empregados em sua resolução e para a obtenção de um resultado mais confiável (CLEMEN; REILLY, 2013).

No caso da proficiência¹ do professor de línguas, a complexidade é ainda maior. Isso porque inúmeras são as competências que devem ser desenvolvidas e mobilizadas por esses profissionais para que o processo de ensino seja bem-sucedido (CONSOLO; TEIXEIRA DA

¹ O termo “proficiência” empregado neste trabalho relaciona-se estritamente à proficiência linguística, mesmo quando empregado sem o acompanhamento de seu modificador.

SILVA, 2007a; CONSOLO; PORTO, 2011). Dentre os principais problemas relacionados à sua avaliação, encontram-se a delimitação do domínio de uso da língua para finalidades específicas, a autenticidade das tarefas e das amostras elicitadas e a extensão da influência de fatores extralinguísticos, como competência estratégica e habilidades de ensino, nessas amostras (DOUGLAS, 2000, 2001; ELDER, 2001). Do mesmo modo, a interpretação arbitrária e insuficiente de algumas abordagens de ensino de línguas, especialmente da abordagem comunicativa, em uma tentativa de negar a tradicional abordagem gramatical, faz com que professores defendam a prevalência de uma ou outra competência sobre a competência linguística. Assim, o elemento linguístico acaba sendo visto como “vilão” no processo de ensino-aprendizagem de línguas, como se o trabalho com a língua que se ensina fosse elemento de não-criticidade, contradizendo os próprios princípios, a própria concepção de língua da referida abordagem, para a qual a língua é um elemento essencial por meio do qual a comunicação é instaurada (SCHULZ; CUSTÓDIO; VIAPIANA, 2012), e fazendo com que exames de proficiência de línguas sejam evitados e mal interpretados (FULCHER; DAVIDSON, 2007).

Scaramucci (2000) ressalta que a falta de objetivos reais e específicos e de critérios claros de orientação do processo contribuem para reforçar a subjetividade da avaliação e, portanto, torná-la injusta. Scaramucci (2000) e O’Sullivan (2011) destacam, assim, a importância de garantir um processo transparente e participativo, ou seja, que integre, em sua base desenvolvimental, a perspectiva dos candidatos a respeito dos parâmetros contextuais, dos processos cognitivos e dos recursos empregados.

Diante desse cenário, no qual as limitações cognitivas humanas podem interferir de maneira significativa no processamento de informações e impactar a aceitação social de avaliações de proficiência, faz-se necessário buscar abordagens que permitam uma modelagem estruturada dos problemas, ou seja, que considerem as opiniões pessoais do avaliador no processo decisório, uma vez que a sistematicidade do pensamento pode auxiliar o desenvolvimento apropriado do julgamento subjetivo (CLEMEN; REILLY, 2013). Em outras palavras, deve-se buscar uma abordagem que não almeje a eliminação da subjetividade do processo decisório, mas, sim, sua representação de modo compreensível e passível de tratamento.

1.1 A pesquisa

Esta pesquisa insere-se no contexto de investigações sobre a avaliação da proficiência dos professores de línguas realizadas pelo grupo de pesquisa ENAPLE-CCC – Ensino

e Aprendizagem de Língua Estrangeira: Crenças, Construtos e Competências (CNPq/UNESP). Historicamente, as buscas pela definição da proficiência do professor de línguas e pelas estratégias de mensuração de seus componentes iniciaram-se por meio da apresentação, no ano 2000, do TEPOLI – Teste de Proficiência Oral em Língua Inglesa, desenvolvido, exclusivamente para avaliar a proficiência do professor de línguas em sua modalidade oral e suscitado pelo questionamento acerca da definição da proficiência do professor de línguas (CONSOLO, 2004; CONSOLO; TEIXEIRA DA SILVA, 2007b). A primeira aplicação do TEPOLI ocorreu a alunos formandos de um curso de Letras no ano de 2001 e o uso desse teste estendeu-se até o ano de 2007 para os referidos alunos de cursos de licenciatura em Letras e para professores em atuação.

Em 2008, foi proposto o exame EPPLE – Exame de Proficiência para Professores de Língua Estrangeira, que avalia a proficiência do professor de língua estrangeira na língua em que ensina em suas modalidades oral e escrita (CONSOLO, 2008; CONSOLO, 2011a, 2011b). O exame surgiu no formato de entrevista em pares face-a-face, da qual participavam dois examinadores – um responsável pela interação e outro examinador-observador. Em 2011, evoluiu para o formato eletrônico. Neste formato, a interação dá-se exclusivamente com o computador por meio de perguntas e respostas e avaliação passa a ser realizada *a posteriori*, analisando-se a interação gravada por dois avaliadores – na presença de discrepância, um terceiro avaliador é solicitado e, assim, sucessivamente até a obtenção de uma faixa em consenso. A partir de 2012, as aplicações do exame passaram a ser realizadas em ambos os formatos. No entanto, apesar de a abordagem avaliativa do exame ser amplamente utilizada na avaliação de proficiência oral (MAROCHI, 2008), tem-se que ela (i.) é pouco prática, uma vez que o resultado final é dependente da disponibilidade de ao menos duas pessoas para avaliação, e (ii.) apresenta validade de resultados questionável, já que a ausência de pesos e de atributos bem definidos para os critérios acaba por ampliar a subjetividade dos julgamentos, deixando-os sujeitos à interpretação dos avaliadores.

Esta pesquisa dedica-se ao estudo da abordagem avaliativa da proficiência oral de professores de inglês como língua estrangeira empregado por um exame de finalidades específicas, o EPPLE, visando a adequá-lo às exigências contextualmente democráticas impostas pelo momento sócio-histórico e, mais especificamente, por seus usuários (*stakeholders*), garantindo maior transparência e apoio às decisões avaliativas. Assim, tem por objetivo geral reestruturar a abordagem de avaliação e, conseqüentemente, de tomada de decisão para classificação do candidato em faixas de proficiência oral de um exame que tem

por objetivo a avaliação da proficiência em língua estrangeira de professores de línguas. Para tanto, faz uso de métodos de estruturação de problemas e de métodos de apoio à decisão.

Os objetivos específicos são (i) propor uma escala de proficiência oral que incorpore os interesses dos usuários (*stakeholders*) do exame, bem como as características de fala dos candidatos e das tarefas, e que esteja alinhada a indicadores de proficiência linguística internacionais e (ii) propor uma estrutura avaliativa que potencialize a precisão, a transparência, a praticidade e a confiabilidade dos resultados.

Assim, visa responder às seguintes perguntas de pesquisa:

1. Quais critérios devem integrar a avaliação da proficiência linguístico-comunicativo-pedagógica do professor de línguas?
2. Quais critérios da proficiência do professor de línguas são passíveis de mensuração – quantitativa ou qualitativamente?
3. Como mensurar a proficiência linguístico-comunicativo-pedagógica do professor de línguas, garantindo os princípios de validade, confiabilidade e praticidade das avaliações e agregando os valores do grupo de usuários (*stakeholders*) do exame sobre os critérios que a integram?

A escolha pelo trabalho com a avaliação da proficiência linguístico-comunicativo-pedagógica justifica-se, neste estudo, pelo fato de esse conceito de proficiência tratar-se de uma noção recente nos estudos em Linguística Aplicada (CONSOLO, 2011a, 2011b; CONSOLO; TEIXEIRA DA SILVA, 2014, 2016), carecendo, ainda, de maior compreensão. A justificativa estende-se para o fato de a avaliação dessa proficiência, conseqüentemente, impor, por sua natureza, inúmeras dificuldades decisórias. O exame EPPL, por sua vez, foi escolhido por ser o único exame de bases comunicativas que trabalha com a proficiência em língua estrangeira do professor de línguas (CONSOLO, 2011a, 2011b). Já a modalidade oral da linguagem, por sua natureza dinâmica, aberta a níveis maiores de incerteza interpretativa (PLOUGH; BRIGGS; BONN, 2010), além de ser considerada essencial para a efetividade do processo de ensino de línguas (CONSOLO, 2004; CHAMBLESS, 2012).

Além disso, diante do fato de os resultados apresentados pelo EPPL poderem ser utilizados para decidir se um professor está apto ou não a ministrar determinadas aulas, informando-lhe e a seu empregador sobre a qualidade e sobre os limites de sua proficiência, faz-se necessário que esses resultados apresentem altos índices de validade e, por extensão, de confiabilidade. Desse modo, na avaliação de proficiência e, mais especificamente, no EPPL, há a necessidade de se minimizar os impactos da discrepância, ou seja, de adequar a interpretação dos resultados do exame ao seu construto e gerenciar a inerente influência da

subjetividade na tomada de decisão, tornando o processo avaliativo mais estruturado e compreensível. Apesar de ainda não existirem estudos sobre a presença e o gerenciamento das discrepâncias no exame EPPLÉ, especificamente, tal necessidade mostra-se fundamental em virtude do fato de ele consistir em uma avaliação de potencial alto impacto (KANE, 2002; KUNNAN, 2012), desempenhando um papel decisivo na determinação da vida dos candidatos (SCARAMUCCI, 2011). Ademais, apesar de haver técnicas estatísticas clássicas indicadas e reconhecidas para a avaliação dos níveis de concordância entre avaliadores em diferentes momentos e finalidades avaliativas, garantindo, assim, maior confiabilidade ao processo, é imprescindível ressaltar que tais técnicas trabalham com conceitos de diferença e relação, os quais não são necessariamente sinônimos de concordância e podem gerar falsas interpretações (RANGANATHAN; PRAMESH; AGGARWAL, 2017). Do mesmo modo, é fundamental destacar que, por ser um exame em fase de implementação, o exame EPPLÉ costuma contar com uma quantidade de amostras limitada, comprometendo o tamanho da amostra mínimo requerido – igual ou superior a 15 amostras – para a aplicação da maioria dos testes estatísticos paramétricos.

Ademais, diante do caráter atual pouco institucionalizado do EPPLÉ e da pretensão de se tornar um exame de abrangência nacional, o presente estudo traz contribuições ao promover uma oportunidade de remodelagem dos componentes decisórios do exame, propondo uma abordagem mais prática e mais consistente de decisão, e possibilitando sua reavaliação e sua adequação às exigências sociais por transparência e democratização. A modelagem estruturada proposta neste estudo justifica-se, então, como uma tentativa de promover uma maior compreensão do construto, na medida em que engloba elementos que são considerados relevantes por uma ampla gama de usuários (*stakeholders*) do exame. Igualmente, é necessário ressaltar que, por meio da metodologia empregada neste estudo, contribui-se, ainda, para a geração de uma abordagem decisória estruturada que pode servir de apoio inicial à futura elaboração de um sistema automatizado de avaliação. É importante mencionar, contudo, que a abordagem avaliativa proposta por meio deste trabalho não tem a pretensão de eliminar a subjetividade do processo avaliativo, mas, sim, orientar os procedimentos de avaliação dos dados dos examinados no EPPLÉ.

Assim, além da preservação, em termos éticos, da face do candidato, faz-se possível também preservar a face do próprio exame, ampliando sua validade e podendo, inclusive, vir a ampliar sua aceitação. Ressalta-se, por fim, que um melhor entendimento do conceito de proficiência linguístico-comunicativo-pedagógica e de sua avaliação faz-se crucial para cursos

de formação de professores de línguas, possibilitando a criação de estratégias de ensino para promover seu desenvolvimento nos docentes em formação.

1.2 Organização da tese

Este trabalho encontra-se organizado em cinco partes. Esta seção introdutória teve por objetivo apresentar o tema da pesquisa – a saber, a proficiência linguístico-comunicativo-pedagógica do professor de língua estrangeira, especificamente da língua inglesa –, juntamente com o objeto da investigação realizada – a avaliação da proficiência oral do professor de línguas por um exame de finalidades específicas. Diante disso, abordou-se o conceito de avaliar por meio do viés da realização de julgamentos e inferências em paralelo ao conceito de tomada de decisões. A dificuldade e a subjetividade inerentes aos processos decisórios e, por extensão, avaliativos também foram apresentadas, especialmente quando relacionadas a questões de complexidade, incertezas, múltiplos objetivos e decisões em grupo – situações essas que permeiam o ambiente de avaliação de proficiência linguística, especialmente em sua modalidade oral, e que possibilitam questionamentos sobre a confiabilidade desses instrumentos. Na sequência, foram introduzidos os objetivos e as perguntas de pesquisa que orientam o desenvolvimento da investigação, bem como as justificativas para a escolha do objeto de pesquisa.

O primeiro capítulo, referente à fundamentação teórica, visa, como o próprio título sugere, apresentar as teorias que embasam e orientam a investigação. Inicialmente, compõem o capítulo considerações acerca das abordagens de avaliação e de desenvolvimento e operacionalização de avaliações de proficiência linguística ao longo da história, considerando-se suas naturezas, suas finalidades, seus construtos e seus usos, bem como os critérios empregados em avaliações de proficiência consagradas e sugestões da literatura. Na sequência, são discutidos os conceitos e elementos integrantes da proficiência do professor de línguas para ensinar, bem como seus princípios avaliativos. Por fim, é introduzida a noção de análise da decisão, com a apresentação dos métodos de estruturação de problemas e de análise da decisão multicritério, além de considerações acerca das interfaces existentes entre a Linguística Aplicada e a ciência de Teoria da Decisão.

O segundo capítulo, referente à metodologia da investigação, tem por objetivo prover informações a respeito dos contextos e dos participantes da pesquisa, bem como dos instrumentos e procedimentos de coleta, análise e triangulação dos dados. Ressalta-se que, apesar de entender a fundamentação teórica como uma etapa da metodologia da investigação, este trabalho optou por respeitar a ordem canônica de estruturação de trabalhos acadêmicos e

5 CONSIDERAÇÕES FINAIS

Este capítulo tem por finalidade sintetizar os resultados alcançados pela investigação, respondendo, pontualmente, às perguntas de pesquisa apresentadas no capítulo introdutório. A pesquisa dedicou-se ao estudo da abordagem avaliativa da proficiência oral de professores de inglês como língua estrangeira empregado por um exame de finalidades específicas, o EPPL e teve por objetivo geral reestruturar a abordagem de avaliação do candidato em faixas de proficiência oral. Por objetivos específicos, o estudo determinou a proposição de uma escala de proficiência oral que incorporasse os interesses dos usuários (*stakeholders*), bem como as características de fala dos avaliados do exame e que estivesse alinhada a indicadores de proficiência linguística internacionais; e a proposição de uma estrutura avaliativa que potencializasse a precisão, a transparência, a praticidade e a confiabilidade dos resultados.

Para atingir os objetivos propostos, o estudo orientou a investigação por meio de três perguntas de pesquisa. A primeira pergunta tinha por objetivo realizar um levantamento dos critérios que deveriam integrar a avaliação da proficiência do professor de línguas, ou seja, preocupava-se com a definição de critérios que tivessem relevância avaliativa para determinar níveis dessa proficiência. Orientando-se pelos modelos de competência linguística de Bachman (1990) e do CEFR (COE, 2018) e pelos modelos de competência linguístico-comunicativa de Almeida Filho (1993) e do CETF (CAE, 2018), foi realizada a análise da opinião de usuários do exame EPPL, obtida por meio da aplicação de questionários, a esse respeito. Princípios extraídos do método de estruturação de problemas VFT – *Value-Focused Thinking* (KEENEY, 1992, 1994, 1996) orientaram a elaboração dos questionários e a análise dos dados, que resultou em uma hierarquização das preferências.

A investigação voltou-se, então, para a segunda pergunta de pesquisa, que tinha por objetivo determinar, dos critérios elencados por meio da primeira questão, quais eram passíveis, quantitativa ou qualitativamente de mensuração. Das técnicas de mensuração de aspectos relativos à fala em língua estrangeira levantadas por meio de análise da literatura em avaliação de proficiência e presentes no questionário aplicado aos usuários do exame, pôde-se prosseguir para a terceira pergunta, que inquiria a respeito de como mensurar a proficiência linguístico-comunicativo-pedagógica do professor de línguas. Nesta etapa, a pesquisa preocupava-se com a garantia dos princípios de validade, confiabilidade e praticidade das avaliações, bem como

com a agregação dos valores do grupo de usuários do exame sobre os critérios, visando adequar os resultados da avaliação às exigências contextualmente democráticas impostas pelo momento sócio-histórico e, mais especificamente, por seus usuários, garantindo maior transparência e apoio às decisões avaliativas.

Percebeu-se, portanto, a necessidade de se elaborar uma escala de mensuração com finalidade de orientar a avaliação dos desempenhos de fala separada da escala de proficiência com finalidade informativa. Isso porque, verificou-se que os critérios relacionados à organização textual eram dependentes das tarefas do exame, ou seja, eram limitados ao contexto da tarefa e, como tais, não deveriam integrar a escala que tem finalidade inferencial para situações de uso da língua no mundo real. A escala de mensuração foi elaborada por meio da opinião dos usuários, mas também, por meio da criterialidade investigada nos desempenhos potencializados pelas tarefas e nos desempenhos reais dos usuários em cada tarefa. A escala de proficiência foi desenvolvida paralelamente, considerando-se essas variações e os princípios do CEFR (COE, 2018) e CETF (CAE, 2018).

O tratamento da interatividade dos critérios foi realizado por meio da aplicação do método multicritério de análise da decisão *Fuzzy TOPSIS* a aproximadamente 10% dos dados de fala que compõem a investigação. As amostras foram selecionadas por julgamento, ou seja, os candidatos foram selecionados intencionalmente visando à avaliação da capacidade de o método manter avaliações desejáveis e corrigir inconsistências, especialmente com relação ao desempenho dos candidatos nas tarefas de uso da língua para finalidades pedagógicas. Os resultados da reavaliação dos dados de fala revelaram consistência para mais de 80% dos candidatos – havendo, inclusive, a correção de uma avaliação originalmente inconsistente – em relação à escala de proficiência proposta, apontando para a existência de força preditiva positiva por parte dos critérios e dos valores de desempenho que integram a escala de mensuração.

Apesar da necessidade de preparação prévia e, em grande parte, manual dos dados, o que acaba por aumentar o tempo para fornecimento de resultados aos candidatos, é válido destacar que essa preparação garante que a decisão do avaliador seja mais informada, além de garantir uma disponibilidade maior dos dados para pesquisas que alimentem melhorias contínuas no próprio exame. Além disso, a abordagem avaliativa proposta neste trabalho provê níveis maiores de validade dos resultados das avaliações do desempenho de candidatos no exame EPPL ao propiciar maior consideração de variações de desempenho em seu modelo de mensuração e oferecer escalas de avaliação em acordo com as tarefas do exame e com os descritores de proficiência internacionalmente reconhecidos, garantindo, ainda, a minimização de ocorrências de inconsistências graças à informatividade das escalas e à consideração do peso

de preferências de outros decisores, e maior transparência na divulgação dos resultados. Além disso, propicia uma maior compreensão do conceito de proficiência-linguístico-pedagógica graças à análise dos dados de fala dos candidatos e da redefinição dos critérios que o integram.

Deste modo, espera-se que a abordagem de avaliação proposta represente um avanço na integração das opiniões de *stakeholders* do exame e na busca pela (semi-)automatização da correção de dados de fala.

Identifica-se como limitações metodológicas desta pesquisa a quantidade de amostras extraídas do corpus do teste oral do exame EPPL e consideradas na análise deste estudo, as quais, apesar de significativas, impedem a realização de generalizações confiáveis não apenas pelo número total de participantes, mas também por representarem apenas contextos de formação de professores privilegiados; e a quantidade e a variedade alcançada de participantes respondentes dos questionários, que, novamente, impedem a realização de generalizações acuradas e têm suas representatividades concentradas nas regiões Sul e Sudeste do Brasil. Além disso, pode-se questionar a orientação teórico-metodológica dos modos e das unidades de avaliação empregados com relação aos critérios. Isso porque, apesar de as unidades de mensuração utilizadas terem sua eficácia nacional e internacionalmente atestadas na literatura em Linguística Aplicada – como exibido Brown, Iwashita e McNamara (2005), Rimmer (2009), Lim e Galaczi (2010), Manosso (2013), De Jong *et al.* (2013) para citar alguns –, não há consenso a respeito dos melhores indicadores para a proficiência em língua estrangeira de brasileiros.

Desse modo, visando ao aprimoramento das contribuições que podem ser extraídas da realização deste trabalho, sugere-se que a escala de mensuração proposta seja alimentada com uma quantidade maior de dados do corpus do teste oral do exame EPPL de modo a se ampliar a confiabilidade dos indicadores de perfis de desempenho e a capacidade de generalização desses indicadores. Igualmente, sugere-se que as opiniões de um número maior e mais diversificado de usuários do exame seja investigada, evitando-se possíveis enviesamentos avaliativos oriundos de contextos específicos e regionalizados de pesquisas em avaliação. Sugere-se, ainda, conforme apontado pelos resultados da reavaliação dos dados de fala por meio do método multicritério de análise da decisão a revisão dos modelos de tarefas que compõem o teste oral do exame EPPL diante da existência de um foco pronunciado dessas tarefas em situações de uso da língua estrangeira para finalidades gerais em detrimento de situações de uso especificamente pedagógicas, o que faz com que desempenhos de fala pedagógica muito discrepantes da fala geral tenham sua importância reduzida. Sugere-se, por fim, que a aplicação

do método vertex na agregação dos pesos dos decisores de modo a se verificar possíveis diferenças no ordenamento sugerido.

Sugere-se, por fim, o prosseguimento da investigação do processo de validação do exame, especialmente considerado-se a escala de proficiência proposta, segundo os princípios determinados pelo Conselho Britânico (COE, 2009), a saber: a determinação da frequência de variação da classificação dos itens em relação ao CEFR, a determinação de consistência intra-avaliador e o nível de discordância por item.

REFERÊNCIAS

- ALDERSON, J. C. Bands and scores. In: ALDERSON, J. C.; NORTH, B. **Language testing in the 1990s: the communicative legacy**. London: Modern English Publications, British Council, 1991.
- ALMEIDA FILHO, J. C. P. **Dimensões comunicativas no ensino de línguas**. Pontes: Campinas-SP, 1993.
- ALMEIDA FILHO, J. C. P. O professor de língua estrangeira sabe a língua que ensina? A questão da instrumentalização linguística. **Contexturas**, v. 1, n. 1, p. 77-85, 1992.
- ALMEIDA, A. T.; COSTA, A. P. C. S. **Aplicações com métodos de apoio multicritério à decisão**. Recife: Editora Universitária da UFPE, 2003.
- ALMEIDA, S.; MORAIS, D. C.; ALMEIDA, A. T. Agregação de pontos de vista de stakeholders utilizando o Value-Focused Thinking associado a mapeamento cognitivo. **Production**, v. 24, n. 1, p. 144-159, 2014.
- AMIRIAN, S. M. R.; MOQADDAM, H. H.; MOQADDAM, Q. J. Critical Analysis of the models of language proficiency with a focus on communicative models. **Theory and Practice in Language Studies**, v. 7, n. 5, p. 400-407, 2017.
- ANDRELINO, P. J. **Análise da estrutura genérica das instruções na fala do professor de inglês: contribuições para o teste oral do EPPE**. 2014. 359 f. Tese (doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2014.
- ANG-AW, H. T.; GOH, C. C. M. Understanding discrepancies in rater judgement on national-level oral examination tasks. **RELC Journal**, v. 42, n. 1, p. 31-51, 2011.
- ARESE, M. C.; RANGEL, L. A. D.; HALL, J.; ZOTES, L. P.; BONINA, N.; MEIRIÑO, M. J. Aplicação do método TOPSIS na avaliação dos critérios utilizados na seleção de docentes em uma instituição de ensino superior. **Conhecimento e Diversidade**, v. 9, n. 19, p. 47-58, 2017.
- AYRES, M.; AYRES JR., M.; AYRES, D. L.; SANTOS, A. A. S. **Bioestat: aplicações estatísticas nas áreas das ciências biomédicas**. Belém, 2007.
- BACHMAN, L. F. **Fundamental considerations in language testing**. Oxford: OUP, 1990.
- BACHMAN, L. F.; LYNCH, B. K.; MASON, M. Investigating variability in tasks and rater judgments in a performance test of foreign language speaking. **Language Testing**, v. 12, p. 238-257, 1995.
- BACHMAN, L. F.; PALMER, A. S. **Language assessment in practice**. Oxford: OUP, 1996.
- BAFFI-BONVINO, M. A. **Avaliação da proficiência oral em inglês como língua estrangeira: foco na competência lexical e uma proposta para o processo de validação do descritor “vocabulário” de um teste de proficiência para professores de língua inglesa**. 2010.

366 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2010.

BAHABADI, M. D. **Fuzzy TOPSIS Software** – Chen (2000). [S.l.], 2018. Disponível em: https://www.researchgate.net/publication/327468165_fuzzy_topsis_software_-Chen_2000. Acesso em 10 jun. 2019.

BANA E COSTA, C. A. **Structuration, construction et exploitation d'un modèle multicritère d'aide à la décision**. Tese (Doutorado em Engenharia de Sistemas). 1992 – Universidade Técnica de Lisboa, Portugal, 1992.

BARBOSA, S. M. A. D. **Perfis variados de competência linguístico-comunicativa numa LE (inglês) e seu impacto no ensino de línguas**. 2007. 153 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade de Brasília, Instituto de Letras, Brasília, 2007.

BEHZADIAN, M.; OTAGHSARA, S. K.; YAZDANI, M.; IGNATIUS, J. A state-of-the-art survey of TOPSIS applications. **Expert Systems with Applications**, v. 39, p. 13051-13069, 2012.

BORGER, L. **Looking beyond scores: A study of rater orientations and ratings of speaking**. 2014. 177 f. Licenciante thesis (Licenciante degree in Subject Matter Education) – University of Gothenburg, Faculty of Education, Gothenburg, 2014.

BORGES-ALMEIDA, V. **Precisão e complexidade gramatical na avaliação de proficiência oral em inglês do formando em Letras: implicações para a validação de um teste**. 2009. 272 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2009.

BORGES-ALMEIDA, V.; CONSOLO, D. A. Testing oral proficiency: What does pronunciation tell us? **Ilha do Desterro**, n. 60, p. 247-275, 2011.

BRAGA, S. R. P. **Avaliação de proficiência oral nos cursos de Letras: a visão de professores e alunos**. 2013. 167 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade de Brasília, Instituto de Letras, Brasília-DF, 2013.

BROWN, J. D. Open-response items in questionnaires. In: HEIGHAM, J.; CROCKER, R. (ed.). **Qualitative research in Applied Linguistics: A practical introduction**. New York: Palgrave Macmillan, 2009.

BROWN, A. An investigation of the rating process in the IELTS Speaking Module. In: TULLOH, R. (ed.). **Research reports** – Vol. 3. Canberra, Australia: IELTS Australia, 2000.

BOWN, A.; IWASHITA, N.; MCNAMARA, T. An examination of rater orientations and test-taker performance of English-for-Academic-Purposes speaking tasks. **Monograph Series**. ETS, 2005.

BUEDE, D. M. Decision-making and decision analysis. In: GASS, S. I.; FU, M. C. (ed.). **Encyclopedia of operations research and management science**. Boston: Springer, 2013.

- CAE – Cambridge Assessment English. **Framework Components**. Cambridge: CET, 2018.
- CAI, H. Weight-based classification of raters and rater cognition in an EFL speaking test. **Language Assessment Quarterly**, v. 12, p. 262–282, 2015.
- CANALE, M.; SWAIN, M. Theoretical bases of communicative approaches to second language teaching and testing. **Applied Linguistics**, v. 1, n. 1, p. 1-47, 1980.
- CARROLL, J. B. Fundamental considerations in testing for English language proficiency of foreign students. In: ALLEN, H.B. (ed.). **Teaching English as a Second Language**. New York: McGraw Hill, 1961.
- CARVALHO E SILVA, A. L. B. A Análise de necessidades e sua relação com cursos e avaliação de proficiência em inglês para um grupo de pilotos militares brasileiros. **Revista Estudos Linguísticos**, v. 45, n. 2, p. 490-503, 2016.
- CASANAVE, C. P. Case studies. In: PALTRIDGE, B.; PHAKITI, A. **Continuum companion to research methods in Applied Linguistics**. London: Continuum International Publishing Group, 2010.
- CERTO, S. C. Tomada de decisões. In: CERTO, S. C. **Administração moderna**. São Paulo: Pearson, 2005.
- CHAMBLESS, K. S. Teacher's oral proficiency in the target language: Research on its role in language teaching and learning. **Foreign language annals**, v. 45, n. S1, p. S141-S162, 2012.
- CHAPELLE, C. A. Validity argument for language assessment: The framework is simple... **Language Testing**, v. 29, p. 19–27, 2012.
- CHEN, C. T. Extension of the TOPSIS for group decision-making under fuzzy environment. **Fuzzy Sets and Systems**, v. 114, p. 1-9, 2000.
- CHOMSKY, N. **Aspects of the theory of syntax**. Cambridge: MIT Press, 1965.
- CLEMEN, R. T.; REILLY, T. **Making hard decisions with Decision Tools®**. Mason: Cengage Learning, 2013.
- CLAPHAM, C. The Assessment of metalinguistic knowledge. In: ELDER, C. *et al.* **Experimenting with uncertainty: Essays in honor of Alan Davies**. Cambridge: CUP, 2001.
- COE – Council of Europe. **Common European Framework of Reference for Languages: learning, teaching and assessment**. Strasbourg: Language Policy Division, 2001.
- COE – Council of Europe. **Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching and assessment (CEFR)**. Cambridge: Cambridge University Press, 2018.
- COLOMBO, C. S. **O insumo linguístico oral em aulas de inglês como língua estrangeira para crianças: a fala do professor em foco**. 2014. 226 f. Dissertação (Mestrado em Estudos

Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2014.

CONSOLO, D. A. A competência oral de professores de língua estrangeira: A relação teoria-prática no contexto brasileiro. *In*: CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. (ed.). **Olhares sobre competências do professor de língua estrangeira**: da formação ao desempenho profissional. São José do Rio Preto: Editora HN, 2007.

CONSOLO, D. A. A construção de um instrumento de avaliação oral da proficiência do professor de língua estrangeira. **Trabalhos em Linguística Aplicada**, v. 43, n. 2, p. 265-286, 2004.

CONSOLO, D. A. Assessing EFL teachers' oral proficiency: On the development of teacher education programmes and testing policies in Brazil. *In*: CONSOLO, D. A.; GATTOLIN, S. R. B.; TEIXEIRA DA SILVA, V. L. **Perspectivas em avaliação no ensino e na aprendizagem de línguas**: pesquisas e encaminhamentos na formação e na prática docente. Campinas-SP: Pontes, 2017.

CONSOLO, D. A. **Avaliação da proficiência linguístico-comunicativo-pedagógica do professor de línguas**: operacionalização de construto no Exame de Proficiência para Professores de Língua Estrangeira (EPPLE). Projeto trienal de pesquisa – ETAPA I. São José do Rio Preto: UNESP, 2011a.

CONSOLO, D. A. **Avaliação da proficiência linguístico-comunicativo-pedagógica do professor de línguas**: operacionalização de construto no Exame de Proficiência para Professores de Língua Estrangeira (EPPLE). Projeto trienal de pesquisa – ETAPA II. São José do Rio Preto: UNESP, 2011b.

CONSOLO, D. A. **Exame de Proficiência para Professores de Língua Estrangeira (EPPLE)**: definição de construto, tarefas e parâmetros para avaliação em contextos brasileiros. Projeto Trienal de Pesquisa. São José do Rio Preto: UNESP, 2008.

CONSOLO, D. A. Revendo a oralidade no ensino/aprendizagem de línguas estrangeiras. **Revista de Estudos Universitários**, v. 26, n. 1, p. 59-68, 2000.

CONSOLO, D. A.; PORTO, C. F. C. Competências do professor no processo de ensino-aprendizagem de língua estrangeira. **Horizontes de Linguística Aplicada**, a. 10, n. 2, p. 65-86, 2011.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. Em defesa de uma formação linguística de qualidade para professores de línguas estrangeiras: o exame EPPLE. **Horizontes de Linguística Aplicada**, a. 13, n. 1, p. 63-87, 2014.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. Foreign language teacher's proficiency: The implementation of the EPPLE examination in Brazil. *In*: GITSAKI, C.; COOMBE, C. **Current issues in language evaluation, assessment and testing**: Research and practice. Cambridge: Cambridge Scholars Publishing, 2016.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. **Olhares sobre competências do professor de língua estrangeira**: da formação ao desempenho profissional. São José do Rio Preto: Editora HN, 2007a.

CONSOLO, D. A.; TEIXEIRA DA SILVA, V. L. The TEPOLI test: Construct, updated tasks and new parameters to assess EFL teachers' oral proficiency. In: CONGRESSO INTERNACIONAL DA ABRAPUI, 1, 2007, Belo Horizonte. **Anais [...]**. Belo Horizonte: Universidade Federal de Minas Gerais, 2007b. p. 1-12.

CUOGHI, K. G. **Avaliação das alternativas na construção da Usina Hidrelétrica Belo Monte para aplicação de métodos multicritério de análise da decisão**. 2015. 74 f. Dissertação (Mestrado) – Faculdade de Economia, Contabilidade e Administração de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2015.

CUOGHI, K. G.; LEONETI, A. B. A group MCDA method for aiding decision-making of complex problems in public sector: The case of Belo Monte Dam. **Socio-Economic Planning Sciences**, p. 1-29, 2018. DOI: <https://doi.org/10.1016/j.seps.2018.04.002>.

DAMASIO, A. Neuroscience and the emergence of Neuroeconomics. In: GLIMCHER, P. W.; FEHR, E. (ed.). **Neuroeconomics**: Decision making and the brain. Cambridge, Massachusetts: Academic Press, 2014.

DAVIES, A. Ethics, professionalism, rights and codes. In: SHOHAMY, E.; OR, I. G.; MAY, S. **Language Testing and Assessment**. Cham: Springer International Publishing AG, 2017.

DE JONG, N. Fluency in second language testing: insights from different disciplines. **Language Assessment Quarterly**, v. 15, n. 3, p. 237-254, 2018.

DIAS, E. **Avaliação**: meio para intervenção pedagógica na formação do professor de língua estrangeira e (re)construção da sala de aula. 2011. 278 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2011.

DIAS, L.; CLIMACO, J. ELECTRE TRI for groups with imprecise information on parameter values. **Group Decision and Negotiation**, v. 9, n. 5, p. 355-377, 2000.

DOUGLAS, D. **Assessing languages for specific purposes**. Cambridge: CUP, 2000.

DOUGLAS, D. Performance consistency in second language acquisition and language testing research: A conceptual gap. **Second Language Research**, v. 17, n. 4, p. 442-456, 2001.

DOUGLAS, D. **Understanding language testing**. New York: Routledge, 2010.

ECKES, T. Operational rater types in writing assessment: Linking rater cognition to rater behavior. **Language Assessment Quarterly**, v. 9, p. 270-292, 2012.

ECKES, T. Rater types in writing performance assessments: A classification approach to rater variability. **Language Testing**, v. 25, p. 155-185, 2008.

EDEN, C.; SIMS, D. Management science process – Subjectivity in problem identification. **Interface**, v. 11, n. 1, p. 68-74, 1979.

ELDER, C. Assessing the language proficiency of teachers: Are there any border controls? **Language Testing**, v. 2, n. 1, p. 149-170, 2001.

ELDER, C.; MCNAMARA, T.; KIM, H.; PILL, J.; SATO, T. Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. **Language & Communication**, v. 57, p. 14-21, 2017.

ELLIS, R. **Instructed second language acquisition**. Oxford: Basil Blackwell, 1990.

ENSSLIN, L.; MONTIBELLER, G. B.; NORONHA, S. M. **Apoio à decisão: metodologias para estruturação de problemas e avaliação multicritério de alternativas**. Florianópolis: Insular, 2001.

FARHADY, H. The disjunctive fallacy between discrete-point and integrative tests. **TESOL Quarterly**, v. 13, n. 3, p. 347-357, 1979.

FARRELL, T. S.; RICHARDS, J. Teachers' language proficiency. In: FARRELL, T. S. (ed.). **Reflective language teaching: From research to practice**. London: Continuum, 2007.

FERNANDES, A. M. **A (meta)linguagem para explicação gramatical em língua inglesa subsídios para elaboração de tarefas do teste oral do EPPLE (Exame de Proficiência para Professores de Línguas Estrangeiras)**. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências Exatas, São José do Rio Preto, 2016.

FISCHER, A.; GREIFF, S.; FUNKE, J. The process of solving complex problems. **The Journal of Problem Solving**, v. 4, n. 1, p. 19-42, 2012.

FRANCO, L. A.; MONTIBELLER, G. **Problem structuring for multi-criteria decision analysis interventions**. London: The London School of Economics and Political Sciences, 2009.

FULCHER, G. **Testing second language speaking**. New York: Routledge, 2003.

FULCHER, G.; DAVIDSON, F. **Language testing and assessment: An advanced resource book**. New York: Routledge, 2007.

GARCÍA-CASCALES, M. S.; LAMATA, M. T. On Rank Reversal and TOPSIS Method. **Mathematical and Computer Modelling**, v. 56, p. 123-132, 2012.

GENESEE, F. Evaluation. In: CARTER, R.; NUNAN, D. (ed.). **The Cambridge guide to teaching english to speakers of other languages**. Cambridge: Cambridge University Press, 2001.

GLIMCHER, P. W. **Decisions, uncertainty and the brain: The science of Neuroeconomics**. Cambridge, Massachusetts: The MIT Press, 2003.

GLIMCHER, P. W. Neuroeconomics: History. *In*: SQUIRE, L. R. **Encyclopedia of Neuroscience**. Cambridge, Massachusetts: Academic Press, 2009.

GLIMCHER, P. W.; FEHR, E. Introduction: A brief history of neuroeconomics. *In*: GLIMCHER, P. W.; FEHR, E. (ed.). **Neuroeconomics**: Decision making and the brain. Cambridge, Massachusetts: Academic Press, 2014.

GOMES, L. F. A. M.; GOMES, C. F. S.; ALMEIDA, A. T. de. **Tomada de decisão gerencial**: o enfoque multicritério. Rio de Janeiro: Ed. Atlas, 2009.

GRANFELDT, J.; ÅGREN, M. SLA developmental stages and teachers' assessment of written French: Exploring Direkt Profil as a diagnostic assessment tool. **Language Testing**, v. 31, n. 3, p. 285-305, 2014.

GROSS, C. G. Early history of Neuroscience. *In*: ADELMAN, G. (ed.). **Encyclopedia of Neuroscience**. Basel: Birkhäuser, 1987.

GROSS, C. G. History of Neuroscience: Early Neuroscience. *In*: SQUIRE, L. R. **Encyclopedia of Neuroscience**. Cambridge, Massachusetts: Academic Press, 2009.

GUIBOUT, A.; MARTEL, J. M. Tentative guidelines to help choosing an appropriate MCDA method. **European Journal of Operational Research**, v. 109, p. 501-521, 1998.

HAMP-LYONS, L. Holistic writing assessment of LEP students. *In*: OFFICE of Bilingual Education and Minority Languages Affairs (ed.). **Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues**: Focus on evaluation and measurement, Vol. 2. Washington, D.C.: US Government Office of Bilingual Education and Minority Languages Affairs, 1992.

HAWKINS, J. A.; BUTTERY, P. Using learner language from corpora to profile levels of proficiency: Insights from the English Profile Programme. *In*: TAYLOR, L.; WEIR, C. J. (ed.). **Language Testing Matters**: Investigating the Wider Social and Educational Impact of Assessment. Cambridge: CUP, 2009.

HAWKINS, J. A.; BUTTERY, P. Criterial features in learner corpora: Theory and illustrations. **English Profile Journal**, v. 1, n. 1, p. 1-23, 2010.

HOUT, R.; VERMEER, A. Comparing measures of lexical richness. *In*: DALLER, H.; MILTON, J.; TREFFERS-DALLER, J. (ed.). **Modelling and assessing vocabulary knowledge**. Amsterdam: Benjamins, 2007.

HYMES, D. Two types of linguistic relativity. *In*: BRIGHT, W. **Sociolinguistics**. The Hague: Mouton, 1966.

HUHTA, A.; ALANEN, R.; TARNANEN, M.; MARTIN, M.; HIRVELÄ, T. Assessing learners' writing skills in a SLA study: Validating the rating process across tasks, scales and languages. **Language Testing**, v. 31, n. 3, p. 307-328, 2014.

IWASHITA, N.; BROWN, A.; McNAMARA, T.; O'HAGAN, S. Assessing levels of second language speaking proficiency: How distinct? **Applied Linguistics**, v. 29, n. 1, p. 24-49, 2008.

JACOBY, S.; MCNAMARA, T. Locating competence. **English for Specific Purposes**, v. 18, n. 3, p. 213-241, 1999.

JAMIESON, J.; POONPON, K. Developing analytic rating guides for TOEFL iBT integrated speaking tasks. **Research Report**. ETS, jul. 2013.

JENKINS, J. **The phonology of English as an International Language**. Oxford: Oxford University Press, 2000.

JIN, T.; MARK, B. Distinguishing features in scoring L2 Chinese speaking performance: How do they work? **Language Testing**, v. 30, n. 1, p. 23-47, 2012.

JOERIN, F.; THERIAULT, M.; MUSY, A. Using GIS and outranking multicriteria analysis for land-use suitability assessment. **International Journal of Geographical Information Science**, v. 15, n. 2, p. 153-174, 2001.

KANE, M. An argument-based approach to validation. **Psychological Bulletin**, v. 112, p. 527-535, 1992.

KANE, M. Certification testing as an illustration of argument-based validation. **Measurement: Interdisciplinary Research and Perspective**, v. 2, n. 3, p. 135-170, 2004.

KANE, M. Validating high-stakes testing programs. **Educational Measurement: Issues and Practices**, v. 21, n. 1, p. 31-, 2002.

KEENEY, R. L. Creativity in decision-making with Value-Focused Thinking. **Sloan Management Review**, v. 35, n. 4, p. 33-41, 1994.

KEENEY, R. L. **Value-Focused Thinking: A path to creative decision-making**. Cambridge: Harvard University Press, 1992.

KEENEY, R. L. Value-Focused Thinking: Identifying decision opportunities and creating alternatives. **European Journal of Operational Research**, v. 92, p. 537-548, 1996.

KEENEY, R. L.; MCDANIELS, T. L. Value-Focused Thinking about strategic decisions at BC Hydro. **Interfaces**, v. 22, n. 6, p. 94-109, 1992.

KEENEY, G. L.; VON WINTERFELD, D. Identifying and structuring the objectives of terrorists. **Risk Analysis**, v. 30, n. 12, p. 1803-1816, 2009.

KOL, J. P. T. **Avaliando as competências dos professores de língua estrangeira (inglês): um estudo sobre a validade de uma certificação internacional**. 2009. 296 f. Dissertação (Mestrado em Linguística Aplicada) - Universidade de Brasília, Instituto de Letras, Brasília, 2009

KROHLING, R. A.; CAMPANHARO, V. C. Fuzzy TOPSIS for group decision making: A case study for accidents with oil spill in the sea. **Expert Systems with Applications**, v. 38, p. 4190-4197, 2011.

KUIKEN, F.; VEDDER, I. Rater's decisions, rating procedures and rating scales. **Language Testing**, v. 31, n. 3, p. 279-284, 2014.

KUNNAN, A. J. High-Stakes Language Testing. In: CHAPELLE, C. (ed.). **The encyclopedia of Applied Linguistics**. Malden: Wiley, 2012.

LIMA JÚNIOR, F. R. **Comparação entre os métodos Fuzzy TOPSIS e Fuzzy AHP no apoio à tomada de decisão para seleção de fornecedores**. 2013. 150 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, 2013.

LLURDA, E. On competence, proficiency, and communicative language ability. **International Journal of Applied Linguistics**, v. 10, n. 1, p. 85-96, 2000.

LUOMA, S. **Assessing speaking**. Cambridge: CUP, 2004.

MAROCHI, T. B. Testes orais individuais e em pares: um estudo das diferenças de desempenho. **DELTA**, v. 24, n. 1, p. 23-49, 2008.

MAROSSO, P. Z. **O emprego do acento nuclear em língua inglesa por falantes brasileiros: um estudo comparativo na era do Inglês como língua franca**. 2013. 63 f. Trabalho De Conclusão de Curso (Graduação em Licenciatura em Letras) – Universidade Tecnológica Federal do Paraná, Curitiba, 2013.

MCNAMARA, T. Assessment of second language proficiency. In: BROWN, K. (Ed.). **Encyclopedia of language and linguistics**. Amsterdam: Elsevier, 2006.

MCNAMARA, T. **Language testing**. Oxford: Oxford University Press, 2000.

MCNAMARA, T. Language testing. In: DAVIES, A.; ELDER, C. **The handbook of Applied Linguistics**. Oxford: Blackwell, 2004.

MCNAMARA, T. **Measuring second language performance**. London: Longman, 1996.

MCNAMARA, T.; ROEVER, C. **Language testing: The social dimension**. Oxford: John Wiley & Sons, 2006.

MCWHITNEY, B. **Tools for analyzing talk: Part 2 – The CLAN Program**. Carnegie Mellon: Carnegie Mellon University, 2019.

MESSICK, S. The interplay of evidence and consequences in the validation of performance assessments. **Educational Researcher**, v. 23, n. 2, p. 13-23, 1994.

MESSICK, S. Validity. In: LINN, R. L. (ed.). **Educational measurement**. New York: Macmillan, 1989.

MESSICK, S. **Validity and washback in language testing**. New Jersey: ETS, 1996.

MINGERS, J.; ROSENHEAD, J. Problem structuring methods in action. **European Journal of Operational Research**, v. 152, p. 530-554, 2004.

MIRANDA, C. M. G.; ALMEIDA, A. T. de. Avaliação de pós-graduação com método ELECTRE TRI – o caso de Engenharias III da CAPES. **Revista Produção**, v. 13, n. 3, p. 101-112, 2003.

MORAIS, D. C.; ALENCAR, L. H.; COSTA, A. P. C. S.; KEENEY, R. L. Using Value-Focused Thinking in Brazil. **Pesquisa Operacional**, v. 33, n. 1, p. 73-88, 2013.

MORROW, C. K. Communicative language testing. In: LIONTAS, J. I. **The TESOL Encyclopedia of English Language Teaching**. New York: John Wiley & Sons, 2018.

MUNDA, G. Multiple-criteria decision aid: Some epistemological considerations. **Journal of Multi-Criteria Decision Analysis**, v. 2, n. 1, p. 41-55, 1993.

MYFORD, C. M.; WOLFE, E. W. **Monitoring sources of variability within the Test of Spoken English assessment system**. New Jersey: ETS, 2000.

NICHOLLS, D. The Cambridge Learner Corpus – error coding and analysis for lexicography and ELT. In: ARCHER, D.; RAYSON, P.; WILSON, A.; MCENERY, T. (ed.). **Proceedings of the Corpus Linguistics**, 2003, Conference, UCREL technical paper number 16, UCREL, Lancaster University, 2003.

NORTH, B. The Common European Framework of Reference: development, theoretical and practical issues. In: A NEW DIRECTION IN FOREIGN LANGUAGE EDUCATION: The Potential of the Common European Framework of Reference for Languages, 2006. **Proceedings [...]**. Japan: Osaka University of Foreign Studies, mar. 2006.

NORTH, B.; SCHNEIDER, G. Scaling descriptors for language proficiency scales. **Language Testing**, v. 5, n. 2, p. 217-262, 1998.

O'SULLIVAN, B. Language testing. In: SIMPSON, J. (ed.). **Routledge handbook of Applied Linguistics**. Oxford: Routledge, 2011.

PLOUGH, I. C.; BRIGGS, S. L.; BONN, S. V. A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. **Language Testing**, v. 27, n. 2, p. 235-260, 2010.

POLLITT, A.; MURRAY, N. L. What raters really pay attention to. In: MILANOVIC, M.; SAVILLE, N. (ed.). **Language Testing 3: Performance, testing, cognition and assessment**. Cambridge: Cambridge University Press, 1996.

RAJAGOPALAN, K. 'World English' or 'World Englishes'? Does it make any difference? **International Journal of Applied Linguistics**, v. 22, n. 3, p. 374-391, 2012.

RAJAGOPALAN, K. O inglês como língua internacional na prática docente. In: LIMA, D. C. (org.). **Ensino e aprendizagem de língua inglesa**. São Paulo: Editora Parábola, 2009.

RANGANATHAN, P.; PRAMESH, C. S.; AGARWAL, R. Common pitfalls in statistical analysis: Measures of agreement. **Perspectives in Clinical Research**, v. 8, p. 187-191, 2017.

REA-DICKINS, P. Fossilization or evolution: The case of grammar testing. In: ELDER, C. *et al.* **Experimenting with uncertainty**: essays in honor of Alan Davies. Cambridge: CUP, 2001.

RIAZI, A. M. **The Routledge encyclopedia of research methods in Applied Linguistics**: Quantitative, qualitative and mixed-methods research. New York: Routledge, 2016.

RICHARDS, H. M.; CONWAY, C.; ROSKVIST, A.; HARVEY, S. Foreign language teachers' language proficiency and their language teaching practice. **The Language Learning Journal**, v. 41, n. 2, p. 231-246, 2013.

RIMMER, W. Can what counts in complexity be counted? **Language Studies Working Papers**, v. 1, p. 25-34, 2009.

ROBBINS, S.; JUDGE, T.; SOBRAL, F. **Comportamento organizacional**: teoria e prática no contexto brasileiro. São Paulo: Pearson, 2010.

ROSENHEAD, J. What's the problem? An introduction to problem structuring methods. **Interfaces**, v. 26, n. 6, p. 117-131, 1996.

ROSENHEAD, J.; MINGERS, J. (ed.). **Rational analysis for a problematic world revisited**. Chichester: Wiley, 2001.

RUTISCHINI, A. Neuroeconomics: What have we found, and what should we search for. **Current Opinion in Neurobiology**, v. 19, n. 6, p. 672-677, 2009.

SAWAKI, Y. Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. **Language Testing**, v. 24, p. 355-390, 2007.

SCARAMUCCI, M. V. R. Proficiência em LE: considerações terminológicas e conceituais. **Trabalhos em Linguística Aplicada**, v. 36, p. 11-22, 2000.

SCARAMUCCI, M. V. R. Validade e consequências sociais das avaliações em contextos de ensino de línguas. **Linguarum Arena**, v. 2, p. 103-120, 2011.

SCHOFFEN, J. R. **Gêneros do discurso e parâmetros de avaliação de proficiência em português como língua estrangeira no exame CELPE-BRAS**. 2009. 192 f. Tese (Doutorado em Linguística Aplicada) – Universidade Federal do Rio Grande do Sul, Instituto de Letras, Porto Alegre, 2009.

SCHULZ, L. O.; CUSTÓDIO, M. M. C.; VIAPIANA, S. Concepções de língua, linguagem, ensino, aprendizagem e suas repercussões na sala de aula de língua estrangeira. **Pensar Línguas Estrangeiras**, v. 1, n. 1, p. 1-21, 2012.

SHEPARD, L.A. Evaluating test validity. **Review of Research in Education**, v. 19, p. 405-450, 1993.

SHOHAMY, E. **Language policy**: Hidden agendas and new approaches. London: Routledge, 2006.

SHOHAMY, E. Language tests as language policy tools. **Assessment in Education**, v. 14, n. 1, p. 117-130, 2007.

SHOHAMY, E. **The power of tests**. London: Longman, 2001.

SHOHAMY, E.; OR, I. G. Volume editors' introduction to "Language Testing and Assessment". In: SHOHAMY, E.; OR, I. G.; MAY, S. **Language testing and assessment**. Cham: Springer International Publishing AG, 2017.

SIEBERT, J.; VON WINTERFELD, D.; JOHN, R. S. Identifying and structuring the objectives of the "Islamic State of Iraq and the Levant" (ISIL) and its followers. **Decision Analysis**, v. 13, n. 1, p. 26-50, 2016.

SIMON, H. A. **Comportamento administrativo**: estudo dos processos decisórios nas organizações administrativas. Rio de Janeiro: Fundação Getúlio Vargas, 1965.

SMITH, J. K. Reconsidering reliability in classroom assessment and grading. **Educational Measurement: Issues and Practices**, v. 22, n. 4, p. 26-33, 2003.

SMITH, D. V.; HUETTEL, S. A. Decision neuroscience: Neuroeconomics. **WIREs Cognitive Science**, v. 1, p. 854-871, 2010.

SPOLSKY, B. History of language testing. In: SHOHAMY, E.; OR, I. G.; MAY, S. **Language testing and assessment**. Cham: Springer International Publishing AG, 2017.

STASSENKO, I.; SHOPINSKAJA, L.; LIIV, S. Investigating cultural variability in rater judgements of oral proficiency interviews. **Eesti Rakenduslingvistika Ühingu Aastaraamat**, v. 10, p. 269-281, 2010.

TAGUCHI, N. Rater variation in the assessment of speech acts. **Pragmatics**, v. 21, n. 3, p. 453-471, 2011.

TASKIN, H.; ÜSTÜN, Ö; DELIKTAS, D. Fuzzy MCDAM approach for oral examination in erasmus student selection process. **Journal of the Institute of Science & Technology of Dumlupinar University**, v. 32, p. 21-40, 2013.

TAYLOR, D. The meaning and using of the term competence in linguistics and applied linguistics. **Applied Linguistics**, v. 9, p. 148-168, 1988.

TAYLOR, L.; JONES, N. Cambridge ESOL exams and the Common European Framework of Reference (CEFR). **Research Notes**, v. 24, p. 2-5, 2006.

TEIXEIRA DA SILVA, V. L. **Fluência oral**: imaginário, construto e realidade num curso de Letras. 2000. 218 f. Tese (Doutorado em Linguística Aplicada) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas-SP, 2000.

THOMSON, R. I. Measurement of accentedness, intelligibility and comprehensibility. In: KANG, O.; GINTHER, A. **Assessment in second language pronunciation**. New York: Routledge, 2018.

TOFFOLI, S. F. L.; SIMON, C. V. B. A utilização do modelo multifacetado de Rasch na análise das influências dos avaliadores sobre as avaliações com itens abertos. **Ensaio: avaliação e políticas públicas em educação**, v. 26, n. 1, p. 1303-1323, 2018.

WALKER, R. **Teaching the Pronunciation of English as a Lingua Franca**. Oxford: Oxford University Press, 2010.

WĄTRÓBSKI, J.; JANKOWSKI, J.; ZIEMBA, P.; KARZMARCZYK, A.; ZIOŁO, M. Generalised framework for multi-criteria method selection. **Omega**, p. 1-18, 2018.

WEIR, C. J. **Language testing and validation: An evidence-based approach**. New York: Palgrave Macmillan, 2005a.

WEIR, J. P. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. **Journal of Strength and Conditioning Research**, v. 19, n. 1, p. 231-240, 2005b.

WIDDOWSON, H. Communicative language testing: The art of the possible. In: ELDER, C. *et al.* **Experimenting with uncertainty: Essays in honor of Alan Davies**. Cambridge: CUP, 2001.

XI, X.; SAWAKI, Y. Methods of test validation. In: SHOHAMY, E.; OR, I. G.; MAY, S. **Language testing and assessment**. Cham: Springer International Publishing AG, 2017.

YAN, X. An examination of rater performance on a local oral English proficiency test: A mixed-methods approach. **Language Testing**, v. 31, n. 3, p. 1-27, 2014.

YOUN, S. J. Rater variability across examinees and rating criteria in paired speaking assessment. **Papers in Language Testing and Assessment**, v. 7, n. 1, p. 32-60, 2018.

ZADEH, L. A. Fuzzy set. **Information and Control**, v. 8, n.3, p. 338-353, 1965.

ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning – I. **Information Sciences**, v. 8, p. 199-249, 1975.

ZAVADSKAS, E. K.; MARDANI, A.; TURSKIS, Z.; JUSOH, A.; NOR, K. M. D. Development of TOPSIS method to solve complicated decision-making problems: An overview on developments from 2000 to 2015. **International Journal of Information Technology & Decision-Making**, v. 15, p. 1-38, 2016.

ZOPOUNIDIS, C.; PENTARAKI, K.; DOUMPOS, M. A review of country risk assessment approaches: A new empirical evidence. In: ZOPOUNIDIS, C.; PARDALOS, P. M. **Managing in uncertainty: Theory and practice**. Dordrecht: Springer Science+Business Media B. V., 1998.